

# Máster en Programación avanzada en Python para Big Data, Hacking y Machine Learning

Programación Python para Machine Learning

# LECCIÓN 07

## Lección 7: Support Vector Machines y Naive Bayes.

# ÍNDICE

- ✓ Introducción
- ✓ Objetivos
- ✓ Support Vector Machines (SVM)
  - ✓ Principios
  - ✓ Implementación
- ✓ Naive Bayes
  - ✓ Principios
  - ✓ Implementación
- ✓ Consideraciones a tener en cuenta.
- ✓ Conclusiones

# INTRODUCCIÓN

- ✓ Dos técnicas para clasificación supervisada:
- ✓ SVM: Una de las técnicas más potentes para clasificación (binaria).
- ✓ Naive Bayes: Razonamiento deductivo aplicado a la clasificación.



# OBJETIVOS

Al finalizar esta lección serás capaz de:

- 1 Conocer los principios de las Máquinas de Vectores Soporte (SVM).
- 2 Saber implementar en Python modelos de SVM para resolver problemas de clasificación y regresión.
- 3 Aprender las bases de los modelos Naive Bayes.
- 4 Explorar las herramientas para implementar en Python modelos Naive Bayes para resolver problemas de clasificación.

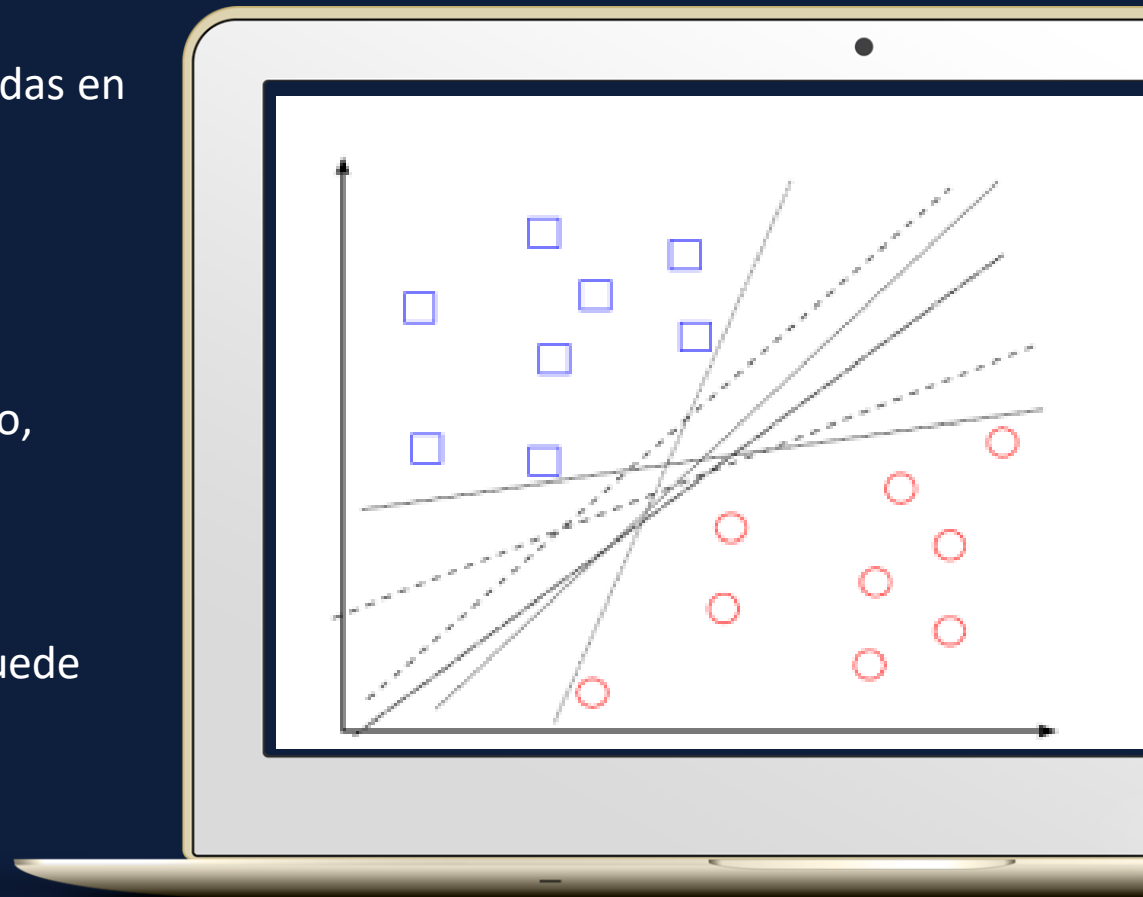
## SUPPORT VECTOR MACHINES (SVM)

Propuestas en los 60, perfeccionadas en los 90.

Modelo supervisado no lineal.

Muy populares por su rendimiento, especialmente, en problemas de clasificación binaria.

¿Cuántas separaciones lineales puede tener un conjunto de instancias?



## Support Vector Machines (SVM)

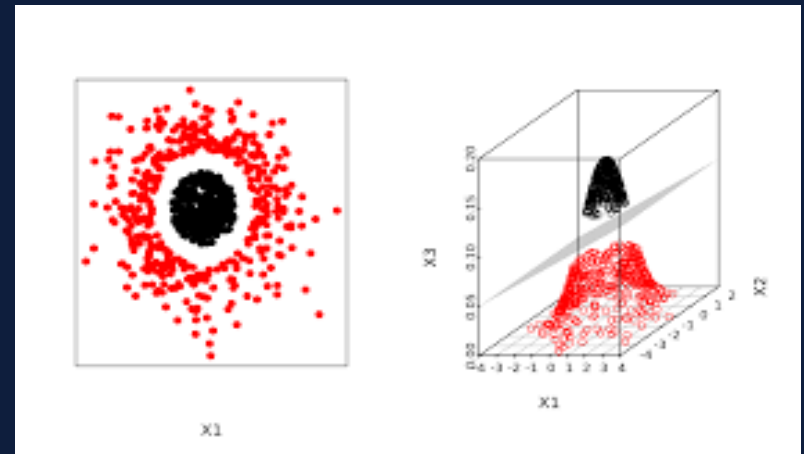
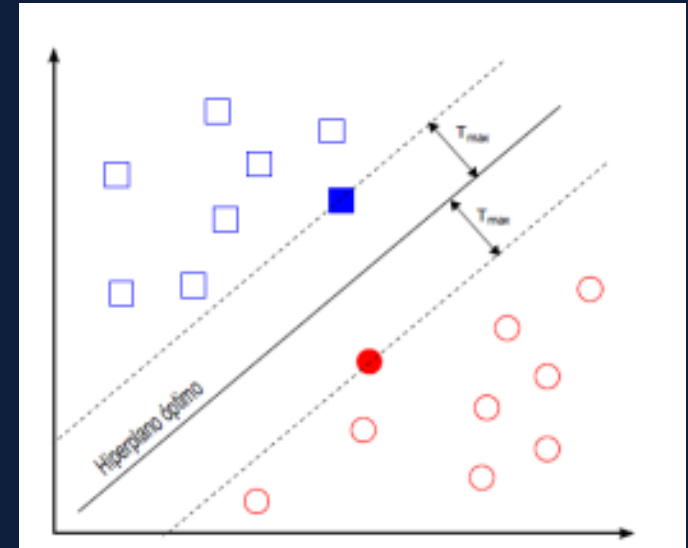
Los SVM tratan de buscar un límite que maximice la distancia de los patrones más cercanos de las dos clases.

Espacio de mayor dimensionalidad, espacio de Hilbert.

Kernel trick.

Peligro: Overfitting del modelo

Por tanto, importancia de parámetros y regularización (C).



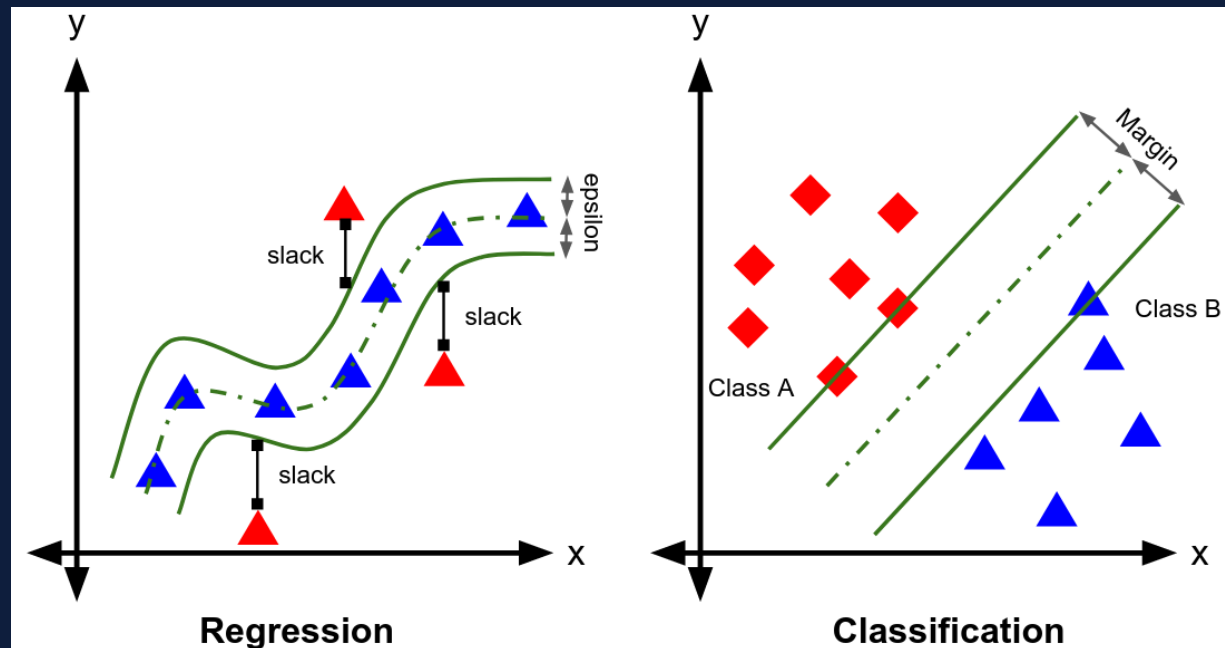
## Support Vector Machines (SVM)

- Función de kernel a utilizar:
  - Kernel Lineal.
  - Kernel Polinomial.
  - Kernel Gausiano.
  - Kernel Sigmoido.
  
- El parámetro C: El parámetro de regularización. Tiene que ser estrictamente positivo.
- El parámetro  $\gamma$ : es el coeficiente del kernel. Solo utilizado en los kernel polinomial, gaussiano y sigmoide.



## Support Vector Machines (SVM)

- Clasificación multiclase: OVO o OVR.
- SVM para regresión.



## NAIVE BAYES

Teorema de Bayes

$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)}$$

$$P(A/B_1, B_2 \dots B_n) = \frac{P(B_1, B_2 \dots B_n/A) \cdot P(A)}{P(B_1, B_2 \dots B_n)}$$

Modelo paramétrico:

- ✓ Se asume independencia absoluta de las variables (ingenuo).
- ✓ Distribución de las variable de entrada: gaussiano, multinomial, Bernoulli...

## NAIVE BAYES

Útil frente a modelos más complejos cuando la cantidad de datos es limitada.

Presenta problemas cuando falta alguna combinación “etiqueta de clase-determinado valor de atributo”.

Clasificación de documentos o procesamiento de textos: El modelo funciona bien en problemas que involucran palabras clave como características, pero no tanto cuando existe una relación fuerte entre los términos o palabras

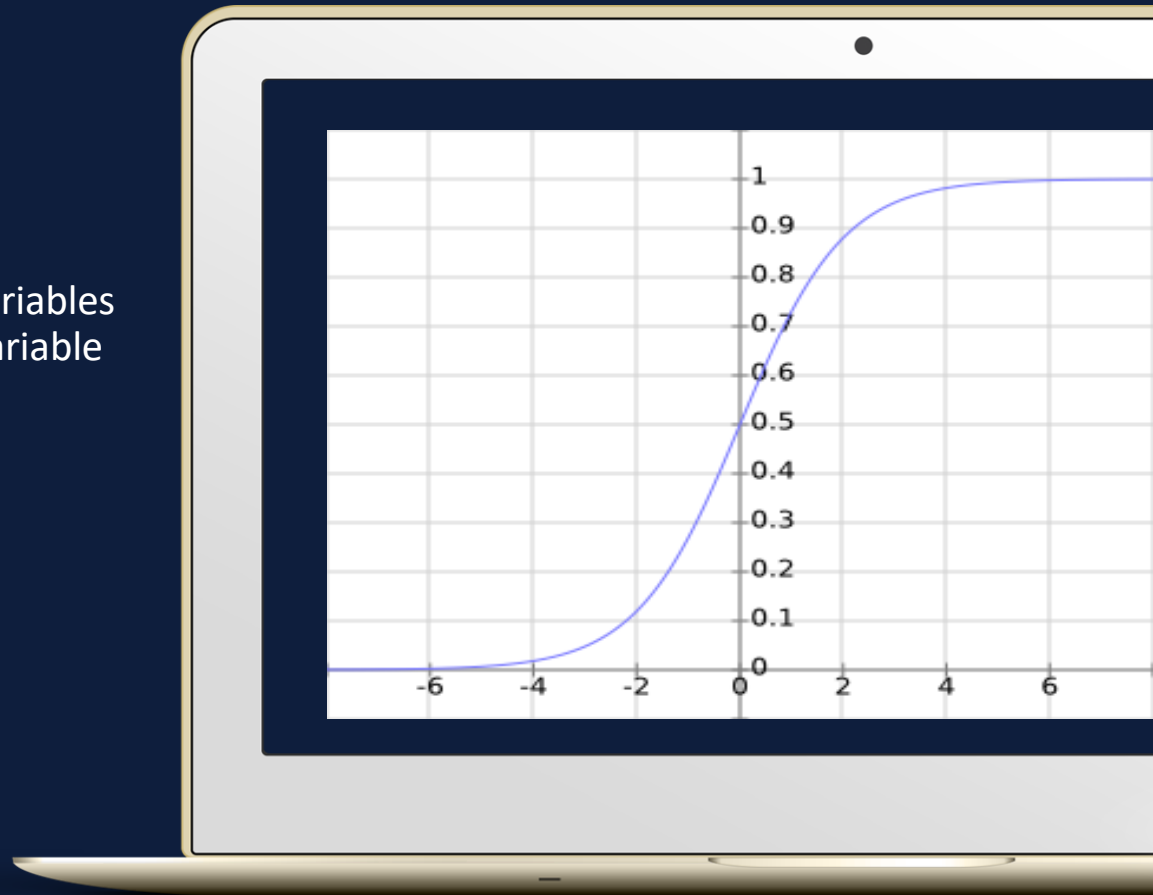
## REGRESIÓN LOGÍSTICA

Modelo supervisado lineal.

Problemas de clasificación.

Trata de modelar mediante una o más variables independientes la probabilidad de una variable dependiente binaria.

$$\text{Sigmoide}(x) = \frac{1}{1 + e^{-x}}$$

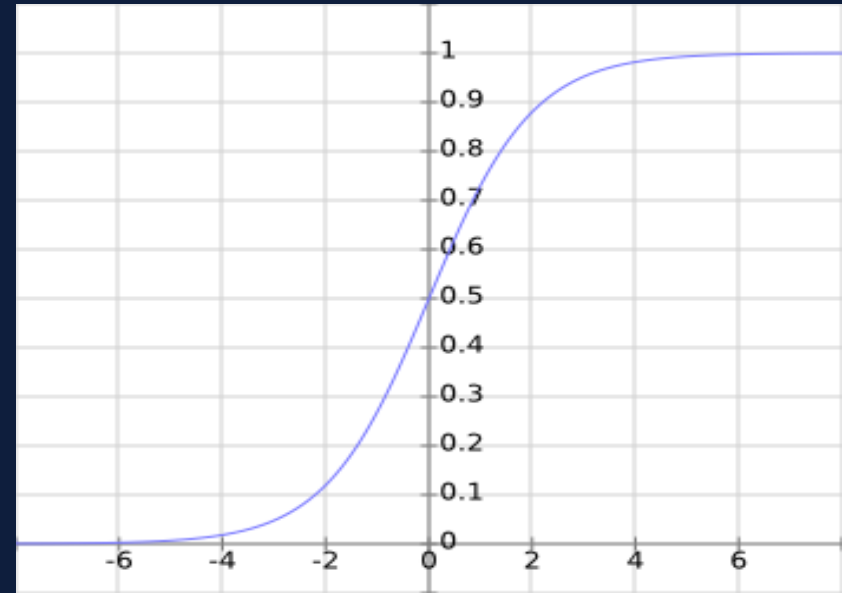


## REGRESIÓN LOGÍSTICA

Regresión logística simple o múltiple.

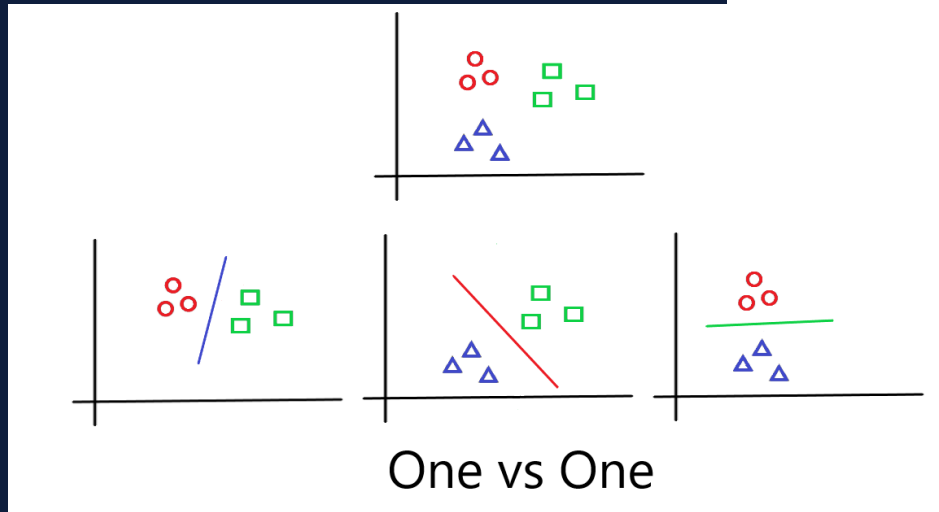
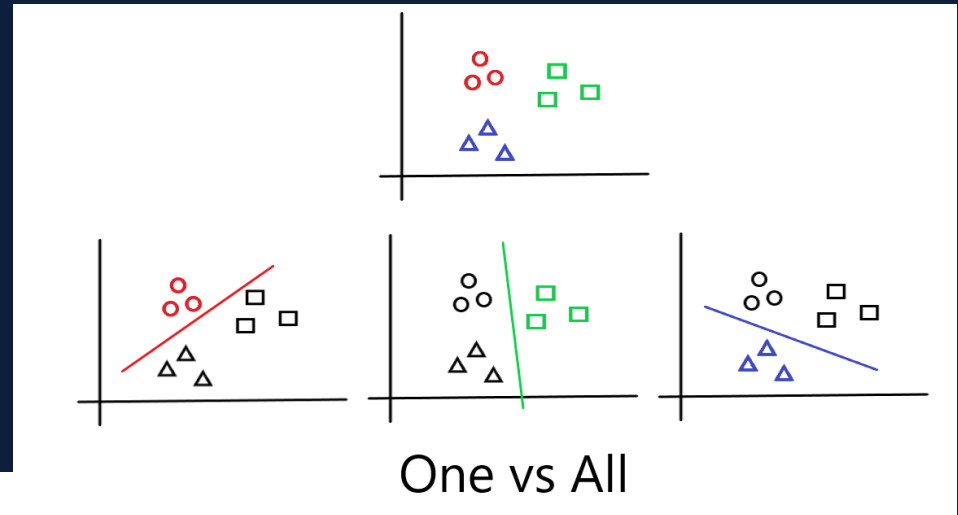
$$P(y = 1 | X = x) = \frac{1}{1 + e^{-(w_1x_1 + w_2x_2 \dots + w_nx_n + w_0)}}$$

Técnica paramétrica.

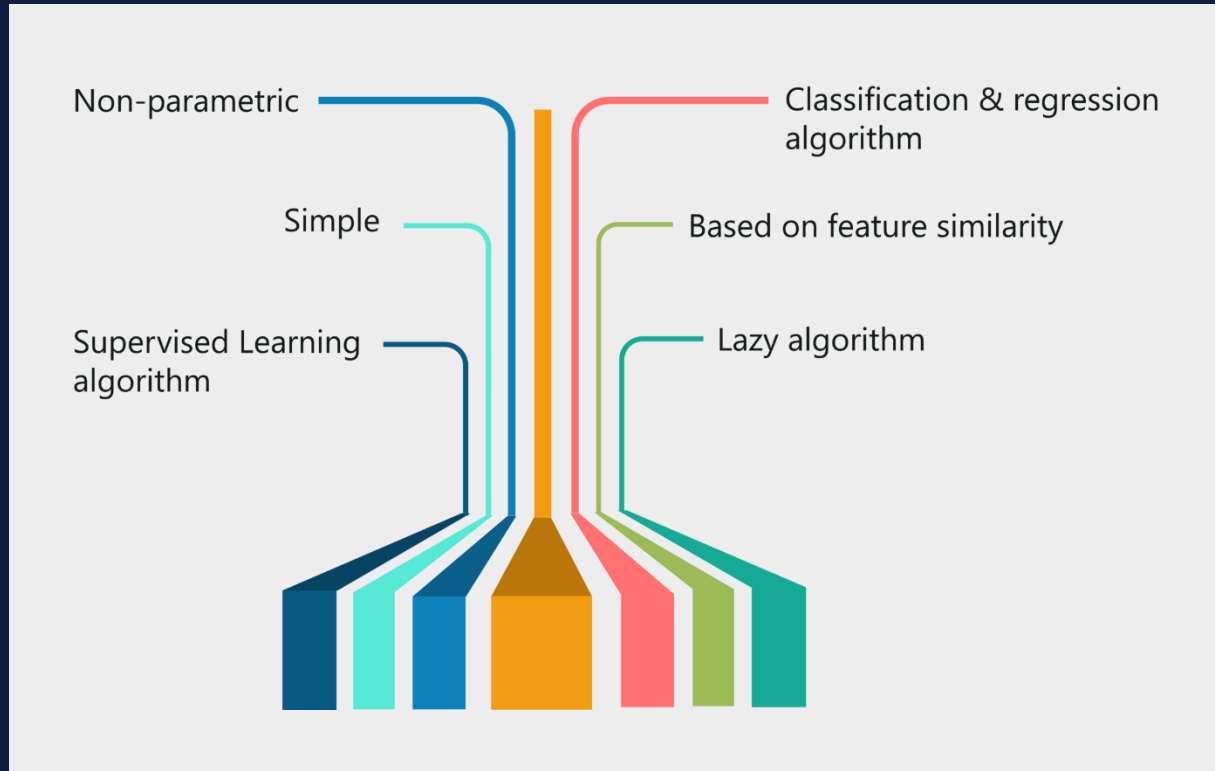


## REGRESIÓN MULTICLASE

- ✓ Estrategia OVA (OVR)
- ✓ Estrategia OVO



## K-Nearest Neighbors

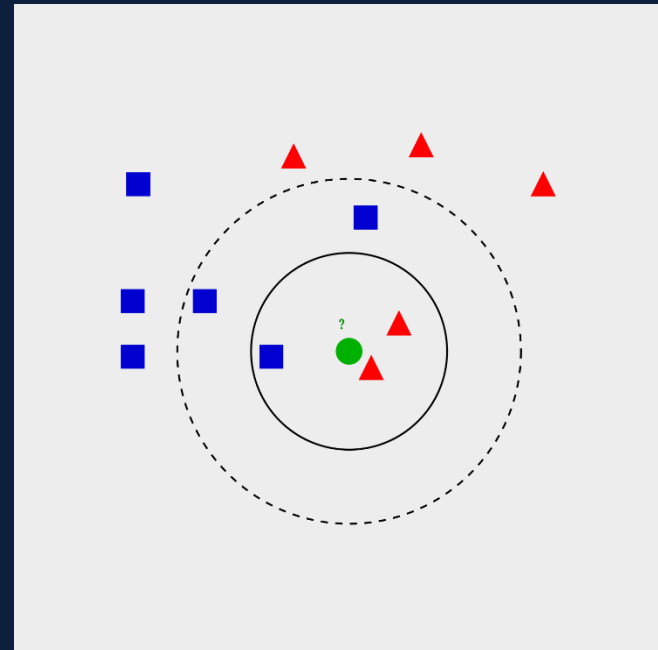


- ✓ Técnica supervisada no lineal.
- ✓ Simple pero muy potente.
- ✓ Perezoso: no entrenamiento

## K-Nearest Neighbors

Fase de predicción:

1. Distancia a todos los puntos.
2. Seleccionar los k patrones más cercanos.
3. Combinar los patrones seleccionados.





MUCHAS GRACIAS POR SU ATENCIÓN



[jperez@grupomainjobs.com](mailto:jperez@grupomainjobs.com)



Javier Pérez Rodríguez  
[www.linkedin.com/in/perezxavi](https://www.linkedin.com/in/perezxavi)



[twitter.com/eiposgrados](https://twitter.com/eiposgrados)



[facebook.com/eiposgrados](https://facebook.com/eiposgrados)



[instagram.com/eiposgrados](https://instagram.com/eiposgrados)